THE IMPACT OF COVID-19 ON ANALYSTS' SENTIMENT ABOUT THE BANKING SECTOR

2021

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2124

Alicia Aguilar and Diego Torres

THE IMPACT OF COVID-19 ON ANALYSTS' SENTIMENT ABOUT THE BANKING SECTOR

THE IMPACT OF COVID-19 ON ANALYSTS' SENTIMENT ABOUT THE BANKING SECTOR (*)

Alicia Aguilar and Diego Torres

BANCO DE ESPAÑA

(*) We acknowledge Ricardo Gimeno, Jesús Ibáñez, Ainhoa Diez-Caballero and Emilio Rodríguez their useful comments received. We also thank participants of the Bank of Spain internal seminar for the feedback provided. Finally, we thank an anonymous referee for his comments on the manuscript.

Documentos de Trabajo. N.º 2124

June 2021

ISSN: 1579-8666 (on line)

Abstract

The use of quantitative tools to analyse the huge amount of qualitative information has been acquiring increasing importance. Market participants and, of course, Central Banks have been involved in this trend. The vast majority of qualitative data can be qualified as non-structured and refers mainly to news, reports or another kind of texts. Its transformation into structured data can improve the availability of information and hence, decision making. This article applies sentiment analysis tools to text data in order to quantify the impact of COVID-19 on the analysts' opinions. Using this methodology, it is possible to transform qualitative non-structured data into a quantitative index that can be used to compare reports from different periods and countries. The results show the pandemic worsens banking sentiment in Europe, which coincides with higher uncertainty in the stock market. There are also regional differences in the decline in sentiment as well as higher divergence is observed across opinions.

Keywords: Sentiment analysis, COVID-19 impact, European banking, analysts' estimates.

JEL classification: G21, C81, D8, C43.

Resumen

La aplicación de herramientas cuantitativas que facilitan el análisis de la inmensa cantidad de información disponible ha ido ganando cada día más importancia. Son varios los participantes del mercado que se han unido a esta tendencia, y los bancos centrales no escapan de ella. Gran parte de la información cualitativa es no estructurada, principalmente en forma de noticias, informes u otro tipo de textos. Por lo tanto, la automatización de este proceso puede incrementar el volumen de información disponible y el proceso de toma de decisiones. Este trabajo se enmarca en esta tendencia, mediante el uso de herramientas de análisis de sentimiento para determinar el impacto del COVID-19 en la opinión de los analistas sobre el sector bancario. Gracias a esta metodología, se logra convertir una información cualitativa, no estructurada, en un índice cuantitativo que permite comparar informes de diferentes períodos y países. Como resultado, se observa un empeoramiento del sentimiento sobre la banca europea, lo que coincide con una mayor incertidumbre en las cotizaciones bursátiles. Además, se aprecian diferencias entre países, así como una mayor divergencia en las opiniones reflejadas en los informes.

Palabras clave: análisis del sentimiento, impacto del COVID-19, bancos europeos, estimaciones de analistas.

Códigos JEL: G21, C81, D8, C43.

1. Introduction

The equity valuation of financial entities is a crucial element for economic and financial markets agents. In that sense, Central Banks play an important role as they supervise the banking sector and monitor risks to financial stability. Valuation indicators can be very diverse, where stock prices, volatility or earnings estimates are among the most commonly tracked. These quantitative indicators provide comparability across time and entities. Additionally, financial analysts and rating agencies provide research and publications that offer their qualitative assessment about different subjects such as rating updates, financial disclosures, questions related to the financial sector, or issues affecting specific entities. The information conveyed in these reports can be very useful as an overview of analysts' opinions and market sentiment during periods of high volatility.

Since the inception of the pandemic, banking stock prices dropped more than general stock indexes (Figure 1), even if institutions such as the European Central Bank (ECB) or the European Banking Association (EBA) stated that banks are now in a better position than in the Great Financial Crisis (GFC). The increasing gap between the banking sector and the general stock indexes has been observed along with a worsening of analysts' outlooks, highlighting prospects of lower profitability and a deterioration of credit quality (see ECB May 2020 Financial Stability Review and EBA 2020*)*, which lead most analysts to revise down earnings per share (EPS) and profitability (ROE) estimates of banks for 2020 (Figure 2).

In that sense, analysts' opinions before and after the inception of the Covid pandemic constitute a useful piece of information about their perspectives for the banking sector, which conveys additional information than the one contained in quantitative indicators. Indeed, several financial providers have created sentiment indicators based on news and research available on their platforms that could help to analyse the impact of the pandemic. Although these indexes give a first approximation, they have some disadvantages. First, they only contain the average sentiment of a sample of reports and not individual values of the index. Secondly, getting the sentiment index for each report could help understand the divergence or disagreement across the pool of opinions. Thirdly, the sentiment index could be biased as they only refer to the opinions from one specific source.

For that reason, the main contribution of this paper is to offer individual indicators of analysts' opinions in order to compare different periods, entities, countries, or reports. This article elaborates this Sentiment Index (IS), and illustrate its usefulness for assessing how analysts' opinions about the banking sector have been downgraded after the inception of Covid-19 and compare its reaction with the one observed in analysts' estimates and in the financial markets.

The remainder of the paper is organized as follows. The next section provides a review of the literature. Section 3 describes how we built the database of reports contained in the analysis, and section 4 defines the methodology used to get the sentiment index (IS). Sections 5 and 6 present the main results and the robustness analysis, respectively. Section 7 compares the IS with other financial indicators and finally, section 8 concludes.

2. Literature review

Text mining techniques applied to financial and economic reports have been of increasing importance for a wide variety of texts such as monetary policy press conferences transcripts, earning calls, or press news. The main objective is gathering qualitative information to evaluate textual tone where the analysis of frequencies of some specific words or topics falls within the most commonly used techniques. Sentiment analysis can be defined as a particular discipline in the field of textual analysis that aims to quantify the tone of a given document through the classification of words into two polarized categories: positive and negative.

Text mining was firstly introduced in 1966 by the researcher Philip J. Stone, who developed the "General Inquirer" (GI) which supposed the creation of the first dictionary (Harvard IV-4) for getting textual tone. This dictionary contains approximately 12.000 words and 77 categories, being "positive", "negative", "weak", "strong", "active" or "passive" the most representative. A dictionary is a collection or list of words classified into some categories. Sentiment analysis is based on the counting of positive and negative words, so the dictionary used is crucial to get the sentiment of a document. Since the creation of the GI, text mining tools have been used in a broad context of text messages and have been readapted to different types of messages and contents.

For example, Tetlock (2007) analyses the daily news media content of the *World Street Journal* to quantify the impact of negative sentiment on financial markets. The paper demonstrates empirically that higher media pessimism can explain lower stock returns. Similarly, Engelberg (2008) constructs an index based on *Dow Jones News Service stories*, which illustrates the number of negative words in the press content using the GI dictionary.

One of the questions that arises in sentiment analysis is whether a dictionary developed in the context of psychology (GI) can be appropriated for financial content. For that reason, Loughran and McDonald (2008) evaluate the tone of 10-K filings[1] of 7852 entities between 1994 and 2008 in the US based on two dictionaries: the Harvard IV-4 and a new negative words classification (LM). The new wordlist has a lower extension but best reflects the financial context as it considers words that appear with a higher frequency in the *SEC filings*. Moreover, the new LM list adds additional categories

[1] 10-K is a comprehensive report filed annually by a publicly-traded company about its financial performance and is required by the U.S. Securities and Exchange Commission (SEC). The report contains much more detail than a company's annual report, which is sent to its shareholders before an annual meeting to elect company directors.

such as *uncertain* or *litigious* and it incorporates words that are most likely used in the financial context but that were not initially included in the Harvard IV-4 dictionary (e.g., “felony”, “litigation”, “restated”, “misstatement”, unanticipated”). The authors find that almost three-fourths of the negative words in the Harvard IV-4 list did not provide a negative tone in financial applications. Furthermore, the sentiment analysis according to the LM classification manages to explain better stock returns after *10-K filings* conference calls.

Henry and Leone (2010) also investigate the question about which dictionary could better reflect financial context. The authors evaluate the textual tone of financial disclosure press conferences based on two types of wordlists. The first one refers to general context dictionaries, such as the GI and the one designed by Roderick Hart (Diction Software, available at http://dictionsoftware.com/diction-overview/’) related to the political context, which classifies words into five categories: Activity, Optimism, Certainty, Realism, and Commonality. Secondly, Henry and Leone (2010) employ their own developed wordlist that was designed for its use in the domain of financial disclosures. The authors defend the use of specific dictionaries[2] to mitigate issues such as polysemy, i.e. words having multiple meanings. For instance, words such as “shares” or “outstanding” are classified as positive in the GI dictionary but their meaning is completely different when applied to the financial context. Henry and Leone (2010) find that financial domain-specific dictionaries outperform GI in measuring the tone of various financial disclosures as they provide higher economic significance for changes in stock returns.

Similarly, McKay et al. (2012) analyse the textual sentiment of financial disclosure press conferences and its impact on the stock market. The authors state that specific dictionaries better reflect the tone of the documents and they employ the HE

[2] The dictionary by Loughran and McDonald (2009) can be classified as a specific dictionary.

dictionary. The work by Engelberg (2008) also defends the use of specific wordlists because Harvard's positive word list may fail to correlate with financial disclosures, due to erroneous classifications.

Feldman et al. (2009) stated that incorporating qualitative information can better explain stock price movements. That way, they measure the textual tone of the Management Discussion and Analysis Sections (MD&A) for a sample of US firms. They construct three sentiment indicators based on the number of positive and negative words as the difference between positive and negative, expressed as a ratio of the total number of words.

Our paper belongs to the set of work that aims to transform qualitative and non-structured information about entities into a quantitative measure of the textual tone that provides a Sentiment Index. More precisely, we apply the two main financial dictionaries, i.e., the one developed by Henry and Leone (2010) and the one by Loughran and McDonald (2009), into analysts' reports about European banks in order to evaluate their opinions. We chose these specific dictionaries because empirical evidence points to a better performance of sentiment indexes generated with specialized dictionaries compared to generalized ones.

The contribution of our work is twofold. First, we use reports from a wide range of sources to provide complete and heterogeneous points of view. Second, the paper offers a comprehensive study of their opinions and compares them at two points in time: before and after the start of the Covid-19 pandemic. For that reason, we also provide a measure of dispersion or discrepancy across analyst's opinions.

3. A description of the database of analysts' reports

We have built a database of analysts' reports which contains documents from 15 European banks, from the five biggest economies of the Euro Area (i.e., Germany, France, Italy, Spain and Netherlands). We have looked for documents referring to specific entities or the whole banking sector. The sample includes 627 specialized reports classified according to the type of source they come from: a) Financial Data Providers (*Bloomberg*); b) Rating Agencies reports (*Moody's*, *Fitch Connect*, *S&P*); and c) Investment Bank Reports (*Deutsche Bank Research*, *Morgan Stanley Research*).

The first group of reports has been obtained from *Bloomberg Intelligence*, where analysts assess the strategy, main risks, and factors affecting the performance of banking entities. These can be defined as technical and specialized reports including ratios, entities' financial results, as well as a detailed analysis of the main drivers the analysts consider when providing their estimates. Frequencies can vary, but the number of publications increases during some specific periods, such as financial results disclosures, dividend calls, or when new monetary policy tools are introduced.

The second group of reports, rating analysts, show the main aspects driving a rating upgrade or downgrade. Their frequency is lower than for the first group, but after the inception of the pandemic, more documents were provided by rating analysts identifying relevant information and key issues that could affect banking entities. For instance, S&P offers the Market Intelligence Tool, where short stories and news are published on a daily basis.

Finally, financial analysts from Investment Banks produce similar documents, sometimes published when estimates are updated. The vast majority of these reports convey information about the principal risks and/or strengths of each entity.

Most of the reports refer to one specific entity and have been obtained for two different periods. However, approximately 30% of Spanish bank documents contain opinions about two or more than two entities while this percentage is lower (13%) in the case of European banks (Table A2 in Annex A).

The first period refers to the two months immediately prior to the start of the Covid - 19 outbreak in Europe (January and February 2020). The second period (post-Covid) let us assess the impact immediately after the beginning of the crisis in April and May. That way, one can analyse the reaction of analysts during a short time window, when the main event observed in the financial markets was the beginning of the pandemic and the implemented lockdowns[3]. Moreover, we excluded March from the analysis given its pronounced downtrend (Figure 1). In fact, the reaction of analysts' estimates was more clearly observed from April onwards (Figure 2).

The sample of Spanish Banks include the five principal listed banks: Santander, BBVA, CaixaBank, Bankia, and Sabadell, which represent around 93% of all banks stock market capitalization (Table A1 in Annex A). Deutsche Bank and Commerzbank[4] constitute the 80% of German Banks, meaning a 80%. In the case of France, we have considered three entities: BNP Paribas, Credit Agricole and Societe Generale which covers 97% of banks stock market capitalization. Moreover, we include the three Italian banks with the highest stock market capitalization, representing approximately 70% of the market: Intesa Sanpaolo, Unicredit, and Mediobanca. Finally, the Dutch sample is constituted by the most relevant entities in the country: ING Bank and ABN Amro, which accounts for 90% of banks stock market. In all analysed countries the weight of each financial entity remains almost stable in the two periods: pre and post-Covid (Figure 3).

4. The methodology to obtain the Index Sentiment

From this original database of analysts' reports, we transform the qualitative content of the reports into numeric values. Specifically, sentiment analysis is based on the

[3] We exclude the evolution afterward as other events such as the measures implemented by country governments and Central Banks, as well as the later recovery.

[4] The rest of the listed entities in this country represent less than 1% of the stock market capitalization and these two entities account for 80% of the market (Table A1 in Annex A).

classification of documents according to two extreme values (positivity and negativity)[5] to get the polarity of each document and in the end, provide a quantitative index. Positive and negative terms can be referred to as connote terms while the rest of the words in a document are defined as neutral.

The Loughran and McDonald (2011) dictionary has been used to define the tone of each word. It contains a list of negative and positive words based on English financial texts. Using this dictionary, one can obtain a Sentiment Index (IS) for each document, and then group them for each country and period (or bank).

The computation of the IS considers connote terms (positive and negative words) as well as neutral words. Following this approach, positivity and negativity indexes (see equations 1 and 2) can be interpreted as a ratio of negative (positive) words over the total, where values range within -1 (all the words in a document are negative) and 1 (all the words are positive).

However, we observe index values that are far from these extreme points, as connote terms represent a relatively low percentages of the total[6]. The IS (equation 3) is computed as the difference between positive and negative words, expressed as a percentage over the total of words in a document. If the value is equal to zero, the sentiment is neutral, whether because the number of negative and positive words coincide or because there are not connote terms[7]. The value of the index conveys information both about the tone (positive or negative) and its magnitude[8]. In that sense, the higher the value of the index (in absolute terms) the more positive or negative the sentiment will be.

Before obtaining the final index, words such as adverbs, prepositions, names and

[5] The positivity (or negativity) of a document is defined as the number of positive (negative) words within the total number of words. The classification of each word is determined by the use of a pre-established sentiment dictionary.
[6] The percentage of connote terms represents approximately 5% of the total number of words in most of the countries.
[7] Words without connotation or neutral words are the ones that can be classified neither as positive nor as negative.
[8] A higher/lower value of the index reflects a higher/lower sentiment.

other terms[9] not offering textual tone have been removed from each document. Moreover, the frequency (number of times a word appears) of each word has been considered for the analysis.

Additionally, the sentiment index accounts for the use of modifiers. The classification based only on negative and positive words can lead to a misinterpretation of the sentiment in some cases. Instead, considering also modifiers that appear near to connote terms can provide a more precise measure. For example, the word "*loss*" connotes negative according to the LM dictionary, but the initial meaning can be altered if it appears together with a "*not*". If modifiers are included in the analysis, this text will be classified as positive, offering a more accurate sentiment.

This methodology modifies the value and, therefore, the sentiment if a positive/negative word appears near to a modified. Concretely, the IS will be computed as expressed in equation 4. The variable $modifier_i$ is defined as a dummy that takes two possible values: -1 if the connote term appears next to a modifier[10] and 1, otherwise.

Following the last example, if the term "*loss*" does not appear next to any modifier, its sentiment will not be changed, i.e., it is considered a negative term, but if, otherwise, a "*not*" is also included, the sentiment will be changed.

The IS considers the number of positive/negative words as well as its frequency. For that reason, the formula described in equation 4 takes into account all repetitions for each modifier and word. For instance, in the previous example, if "loss" appears twice, one time with a modifier and the other without, the final value assigned to the sentence analysed will be neutral.

---

[9] Words such as "basis", "points", "years", "millions", "euros", days and months have not been considered.

[10] We have considered terms with a distance of 4 or less with respect to the positive/negative word.

$$\text{Negativity Index} = \frac{\sum \text{Negative words}}{\sum \text{Total words}} \tag{1}$$

$$\text{Positivity Index} = \frac{\sum \text{Positive words}}{\sum \text{Total words}} \tag{2}$$

$$Sentiment\ Index\ (IS) = \text{Positivity Index} - \text{Negativity Index} \tag{3}$$

$$IS\ modifiers_j = \frac{\left(\left(\sum Postive\ words_i * modifier_i\right) - \left(\sum Negative\ word_i * modifier_i\right)\right)}{\sum Total\ words_j}, \tag{4}$$

where j refers to each report and i identifies each word.

The literature distinguishes two types of indexes: i) the ones that express negativity (or positivity) as the number of negative (positive) words over the total number of words in a document[11] and, ii) the ones considering negativity (or positivity) over the total number of terms with connotation[12] (see more information in Annex B).

In this paper, the analysis is based on the first type of indexes, i.e., the ones computed as a ratio over the total number of words. The sample of documents included in the analysis is obtained from different sources and the length of these documents is heterogeneous so that the number connote terms varies notably across the reports. For that reason, the main advantage is that we can avoid extreme values in the case the number of positive and/or negative word is very low. Moreover, following this approach, the index conveys information both about the tone, and the number of connote terms. Finally, we have checked that correlation between the two type of methods is high and therefore, the conclusions obtained are very similar (see Annex C and Table D2).

[11] See Feldman et al. (2010) or Correa at al. (2018).
[12] See for example: Moreno and González (2020).

5. Impact of Covid-19 on analysts' sentiment about the banking sector

Using the described methodology in the previous section, one can obtain a sentiment index for each document about the tone and its magnitude. Thus, we can compare in a quantitative manner the opinions on each report as a higher (lower) value of the index will reflect a sentiment improvement (deterioration).

The results show lower values of the index during the second period (post-Covid), which suggests a deterioration of analyst's perception about European banks (Figure 4). Moreover, there is a higher frequency of negative values of the IS during the post-Covid period. The value of the IS[13] has decreased in the five analysed countries but one can observe differences across the countries. The highest sentiment "downgrade" can be seen in Italy while the change in analyst's opinions is nearly inexistent in the Netherlands. It is worth mentioning that even before the start of the Covid-19 pandemic, this country showed less favourable opinions.

The impact of Covid has also been reflected in the distribution and dispersion of analysts' opinions (Figure 5). For Spanish banks, a lower disparity is observed during April and May, while in Italy, there was a significant increase in the variety of opinions[14]. The comparison of the index distributions in the two periods let us to observe the most frequent values in each period. Thus, in the pre-Covid period, analyst's opinions about French banks are mostly concentrated on positive values but this trend changed after the pandemic. According to that, the percentage of negative terms increased from 27 to 52% (Figure 6). Finally, after the start of the Covid, analysed reports tend to provide higher level of connotation, as one can perceive from the reduction in the frequency of neutral values of the IS.

---

[13] We consider the median as the main statistic to compare the two periods, as it is more robust to atypical values and in the case of relatively small samples.

[14] See standard deviations before and after the Covid in Table A.3 (Annex A). We test whether the standard deviation is bigger in the post-Covid as compared to pre-Covid through a "Fligner Killeen" test. In Italy, the difference is statistically significant (99%).

Several hypothesis test have been implemented to check whether the deterioration of the sentiment is statistically significant or not. In that sense, the Wilcoxon Rank Sum Test and Quantile Regression are among the principal tools used[15].

The Wilcoxon Rank Sum Test is a non-parametric test and does not require the data to follow a normal distribution. Its main objective is to evaluate if the samples came from two equally-distributed populations[16]. For that reason, this statistical test let us to assess if analysts' sentiment and hence, the IS, is more negative after the Covid or alternatively, if the difference is not statistically significant[17]. The results suggest a significant decline in analysts' sentiment in Spain, Germany, France and Italy, but the change is not significant in the Netherlands (see Table 1 and Annex D).

Quantile Regression can be used to compare different statistics, such as the median or the percentiles of a given distribution. In this article, we refer to the median as the main parameter to compare the values of the IS in the two periods, and we are interested in testing the hypothesis of a median reduction after the Covid. For that purpose, we run a regression for each country, where sentiment index is the objective variable and a dummy variable is added as the independent variable. This dummy takes two possible values: zero during the *pre-Covid* period (January and February) and one in the *post-Covid* period (March and April). That way, the coefficients obtained from the regression will denote the differences in medians between the two periods. The results (see Figure 7 and Table 1) are consistent with the conclusions obtained from the Wilcoxon test and suggest the difference in medians is negative and statistically significant for Spain, France, Germany and, specially, in Italy[18]. In Netherlands the difference in medians is not significant.

---

[15] Please refer to Annex D for additional procedures.

[16] It also refers to the location of the distribution, i.e., whether if the values of the distribution are more concentrated on the positive or negative side. Therefore, this test can be used to assess if there is a location shift.

[17] We use a left-side test instead of a two-sided test, given our objective is determining whether or not there is a shift to the left of the distribution, i.e. if negative values are more frequent.

[18] The difference is statistically significant at 99%.

6. Robustness analysis

This section provides an analysis of the robustness of the conclusions obtained to changes in the sample of reports and to modifications in the dictionary employed to classify the words.

One of the most relevant aspects when computing the IS is the collection of documents we include in the sample. For that reason, we want to ensure the conclusions are not altered if we change slightly the sample of texts evaluated.

Specifically, for each country and period sample, we eliminate randomly a small percentage (5%) of the reports. The procedure has been repeated 100 times, so that we get one-hundred alternative samples for each country and period. Then, the average sentiment is computed for each sample. The results (Table 2) show that, on average, we will get very similar values if we choose randomly one of the alternative samples.

Similarly, the observed shifts in the form and location of each pair of distributions obtained from the alternative samples are independent of the sample we choose (see Figure E1, Annex E). Finally, we check whether if the deterioration of the sentiment is significant or not if we use alternative samples. Thus, for each of the 10,000 combinations[19], we run a Wilcoxon Rank Sum Test and obtain the *p*-values in each case. On average, the *p*-values obtained in each combination will offer the same conclusions

for each country than the ones obtained from the initial sample (see Table 3). Sentiment downgrading is significant in Spain and Germany (at the 90%), being the change stronger in Italy and France, where differences between the two periods are significant for all the sample combinations (see Figure E2 in Annex E).

[19] Considering previous iterations, 100 samples have been created for each country and period. We combine each of the 100 samples in the pre-Covid with each of the 100 samples in the post-Covid, obtaining, 10,000 different combinations.

Additionally, we check the robustness of the results to the nature of the report. As mentioned in section three, the collection of documents included in the analysis is somewhat heterogeneous according to the frequency and type of content. For instance, reports related to earnings release are published regularly on a quarterly basis while others, such as rating opinion are published with a lower regularity in positive times while they increase its frequency in stressed times. Therefore, our proposal is getting the sentiment index distributions for the analysis referred to earnings release, and compare with the overall sample results.

Similar to what we have done previously, the samples are divided according to the period and country, so that the pre-Covid reports will refer to the earnings results of the last quarter of 2019, and the post-Covid covers the results of the first quarter of 2020. That way, one can avoid selection bias such as a potential increase in the number of reports because of negative news as a consequence of the Covid-19.

Figure 8 shows IS distributions including earnings release analysts' reports for each country and period and compares with the distributions with all types of reports. In all countries, the charts reveal a deterioration of the sentiment index also for the earnings release sample. Moreover, the shift seems to be greater for some countries, such as Spain and Germany, while the worsening in analysts' perception is similar in the case of Italy or the Netherlands.

Table 4 shows the results of comparing analyst's sentiment about the banking sector for different type of reports, which are aligned to the ones obtained in the previous section. More precisely, the Wilcoxon Rank Sum Test indicates the change in sentiment is significant in Spain, Italy, France and Germany, but not in Netherlands.

Secondly, we analyse the robustness of the results to the dictionary used for word classification. In that sense, we use an alternative dictionary commonly used in the financial context: Henry and Leone (2020).

The Henry dictionary (HE) has been commonly employed for sentiment analysis in a financial context as it was constructed to determine the textual tone of earnings press releases. The wordlist is formed by 105 positive and 85 negative words[20]. This dictionary reduces the number of connote terms as compared to the Loughran - McDonald (LM) wordlist, which contains 2355 negative and 354 positive words. Thus, the LM dictionary puts more weight on negative terms and hence, we expect using the HE will bring lower connotation and/or less negativity with respect to LM.

The comparison of the two dictionaries confirms the hypothesis that LM provides more negativity to the sentiment index. Moreover, in Germany, we observe a low level of connotation when using the HE dictionary, and it is also the case for France in the post-Covid (Figure 9). For that reason, the LM dictionary employed in the previous section can offer a better approach to analysts' sentiment (Figure 5).

The results indicate changes in analysts' sentiment about the banking sector are robust to the dictionary employed (see table 5 and Figure 9). The values of the IS given by each dictionary suggest sentiment deteriorated after the Covid, as we can observe a shift to the left in the distributions.

7. Relationship between other financial indicators and the IS

The quantitative measure of analysts' sentiment proposed in this article can be compared with other indicators such as earnings or profit estimates as well as stock prices and its volatility. In that sense, one can expect a change in sentiment to coincide with estimate downgrades and/or higher volatility in the stock market. For that reason, we compare the sentiment index (IS) before and after the Covid with: i) Earnings per share estimates (EPS), ii) Return on Equity estimates (ROE), and iii) realized stock prices volatility.

[20] See the list of positive and negative words in Annex F.

Figure 10 shows EPS and ROE estimates for the year 2020 before and after the Covid, suggesting significant drops, while some differences arise between countries. ROE downgrades are deeper in Germany, whose banks already presented the lowest profitability ratios before the inception of the pandemic. Profitability downgrades have lead Spanish banks to lag behind their French peers.

The deterioration of sentiment for European banks coincides with a higher level of uncertainty in the stock market (Figure 11). Thus, the distributions of daily returns during April and May (post-Covid) pointed to higher volatility. On the contrary, the pre-Covid returns show lower dispersion while extreme values are more frequent in April and May. The evolution of the Spanish banking sector differs from the European one, which experienced a better performance (Figure 11).

The quantitative indicators presented in this section point to the heterogeneity of banks' characteristics within and across countries. Therefore, one can think that changes in sentiment index conveyed on analysts' reports might be also driven by economic or fundamental data and not only because of a general pessimistic sentiment caused by the pandemic. For that reason, we explore how analyst's perception has been affected by the evolution of key performance indicators, such as the EPS or ROE ratio.

We approach this question in two ways. The first one consists of exploiting the diversity of analysts' opinions within each entity using the quantile regression methodology. This procedure can be used to identify if the effects of the pandemic or other banks' characteristics are heterogeneous across the entire distribution of the IS, instead of looking only at the median sentiment change. Secondly, we employ fixed effects regressions at the bank level, in order to account for intrinsic attributes of each entity.

Bank characteristics are analysed using EPS values before and after the Covid as well as analysts'estimates for ROE and EPS. We use EPS reported data as its frequency can be matched with the frequency of the IS and provides a key performance ratio of banks, which accounts for both market data (stock prices) and income statement information (quarterly earnings).

Table 6 contains the estimated quantile regressions of the IS of each report using EPS, as the bank explanatory variable. The specification in column 1 is similar to the one described in section 5 to evaluate the difference in medians of the IS before and after the Covid. Therefore, the variable "time dummy" shows the median change in the IS of the entire sample after the inception of the Covid without considering other bank characteristics. The coefficient for the time variable (-0.022) suggests a significant deterioration of analyst's sentiment. In the second column, one can distinguish the effect on the IS driven by changes in fundamental bank data (EPS) from the general deterioration of markets' view after the Covid. The positive and significant value of the EPS coefficient suggests analysts' opinion is also affected by the economic performance of each entity.

Additionally, columns 3 and 4 show this effect is even bigger for reports related to earnings release. This finding is consistent with what one can expect, as EPS are one of the most relevant performance indicators considered in earnings release communications.

The results shown in Table 6 evaluate the effects on the median of the IS. However, one of the advantages of using quantile regressions is evaluating if the effects of a given variable (e.g. EPS) are homogeneous across the different percentiles of the IS. In that sense, Figure 12 shows the effects of the explanatory variables on different deciles of the IS. The positive relationship between EPS and IS can be observed for the central part of the distribution (i.e. for deciles between 20 and 70), which confirms extreme opinions are not necessarily related to fundamental data. Similarly, the effect of time is not significant in the case of extreme values of the IS.

Table 7 shows the analysis of IS changes based on bank fixed effects. The results indicate that inherent bank characteristics account for a high proportion of the variance within the IS (rho=0.83). Moreover, one can observe a significant positive effect of EPS on sentiment analysis, meaning that the higher the decrease (increase) in the value of EPS the higher the deterioration (improvement) of analyst's sentiment. The second and third column show the effect of a change in EPS and ROE analysts' estimates are also significant and positive The coefficient of the constant indicates changes in the IS not driven by the fundamental data included in the regressions. Therefore, it shows that other general factors can explain a significant and negative change in analysts' sentiment, that stays around 0.02. This means that, after controlling for bank characteristics and EPS, the post-Covid mean index is around 0.02 points lower than the pre-Covid value.

We can conclude that both type of approach, i.e. quantile regression and fixed effects illustrate that even controlling for banks' characteristics the pandemic had a significant impact on analysts' sentiment.

8. Conclusions

The sentiment index presented in this article offers a quantitative measure of the analysts and rating agencies' opinions about the European banking sector conveyed in the reports and research publications. Thus, the index constitutes a useful tool to gather market perception. Moreover, through the transformation of qualitative into quantitative data, we are able to compare this information across documents, entities and periods and check if the qualitative opinions are aligned with earnings estimates or stock prices.

The computation of our index provides a unique indicator that reflects the perception of a wide variety of sources and analysts and can be used to assess all the spectrum of opinions. Moreover, we demonstrated that it can be used to evaluate the impact of specific events. We found empirical evidence for significant deterioration of the sentiment about the banking sector after Covid in almost all countries analysed, except for the Netherlands, where the deterioration is not statistically significant. This more pessimistic perspective is aligned with the higher level of uncertainty observed in the stock market and estimates downgrades. Additionally, we found EPS and ROE account for a non-negligible part of sentiment decline and even controlling for these variables, the pandemic had a significant negative impact in analysts' sentiment.

The impact of Covid has also been reflected in the distribution of our index, where there are differences in the dispersion of analyst' sentiment. For most of the European banks, a lower disparity is observed during April and May so that there is a clear consensus about the deterioration of sentiment. In Italy, there was a significant increase in the variety of opinions, perhaps reflecting the uncertainty related to the impact of Covid on the Italian economy, a country that was affected earlier by the pandemic.

Moreover, the results are robust to the use of alternative dictionaries or samples. The dictionary employed in the analysis and the decision to consider modifiers is relevant to determine the level of sentiment index in each period and country. For instance, the dictionary defined by Loughran and McDonald gives more negativity to the index than the one developed by Henry and Leone. Even though, the decision of which wordlist to use does not affect the results that point out a deterioration of analyst's sentiment. Regarding the sample of reports, we have found the impact on sentiment is not biased by the opinions reflected on a specific source or kind of document.

We are aware our analysis has certain limitations. First, the lack of a sufficiently large sample of daily analysts' specialized reports for each entity makes it difficult to construct a higher frequency index that could be used to analyse correlation with other quantitative indicators such as stock prices. One possible approach could be using short daily news of the banking sector available from different sources or social media opinions. Moreover, building up a daily sentiment index could be interesting for monitoring the banking sector. For instance, Central Banks could be interested in using some kind of early warning system to check if the sentiment of a particular entity worsens significantly with respect to their peers.

Going forward, our paper can be the reference for future work on event studies for sentiment analysis. For instance, the approach developed in this article could be expanded to a longer period, entities and/or countries in order to evaluate the impact of different events or the reaction of different sectors to a particular situation. Indeed, it could be interesting looking at the sentiment impact of Covid on distinct sectors or the sentiment reaction to developments related to the virus in each entity or country.

REFERENCES

Apergis, N., and I. Pradigis (2019). “Stock Price Reactions to Wire News from the European Central Bank: Evidence from Changes in the Sentiment Tone and International Market Indexes”*International Advances in Economic Research,* 25, pp. 91-112.

Brown, S., and J. W. Tucker (2111). “Large-Sample Evidence on Firms’ Year-over-Year MD&A Modifications”, *Journal of Accounting Research*, 49, pp. 309-346.

Correa, R., K. Garud, J. Londono-Yarce and N. Mislang (2017). “Constructing a Dictionary for Financial Stability”, *IFDP Notes 2017-06-28,* Board of Governors of the Federal Reserve System (U.S.).

Engelberg, J. (2008). *Costly information processing: Evidence from earnings announcements*, Working paper, Northwestern University.

European Central Bank (2020). *Financial Stability Review* (May).

European Banking Authority (2020). *The EU Banking Sector: First Insights into the COVID-19 impacts,* Thematic Note EBA/REP/2020/17.

Feldman, R., S. Govindaraj., J. Livnat , and B. Segal (2009). “Management’s tone change, post earnings announcement drift and accruals”, *Review of Accounting Studies,* 15, pp. 915-953

Feldman, R., S. Govindaraj, J. Livnat and B. Segal (2008). *The incremental information content of tone change in management discussion and analysis,* Working paper, INSEAD.

Henry, E. (2006). “Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm”, *Journal of Emerging Technologies in Accounting*, 3, pp. 1-19.

Henry, E. (2008). “Are Investors Influenced by How Earnings Press Releases are Written?”, *Journal of Business Communication*, 45, pp. 363-407.

Henry, E., and A. J. Leone (2010). “Measuring qualitative information in capital markets research”, *Research Methods & Methodology in Accounting eJournal*.

Li, F. (2008). “Annual report readability, current earnings, and earnings persistence”, *Journal of Accounting and Economics*, 45, pp. 221-247.

Loughran, T., and B. McDonald (2008). *When is a liability not a liability?,* Working Paper, University of Notre Dame.

Loughran, T., and B. McDonald (2011). “When is a liability not a liability? Textual analysis”, *Journal of Finance*, 66, pp. 35-65.

Moreno, A., and C. González (2020). *Análisis del sentimiento del Informe de Estabilidad Financiera*. Documentos de Trabajo, nº 2011, Banco de España.

McKay, P., S. Doran, J. Peterson, R. David, Bliss, and A. Barbara (2012). “Earnings Conference Calls and Stock Returns: The incremental informativeness of Textual Tone”, *Journal of Banking and Finance*, 36(4), pp. 992-1011.

Tetlock, P. C. (2007). “Giving content to investor sentiment: The role of media in the stock market”, *Journal of Finance*, 62, pp. 1139-1168.

Table 1: Comparison of statistical hypothesis tests

	Quantile regression		Wilcoxon Rank Sum Test (left-side)	
	test	p-value	test	p- value
Spain	-0.0166**	0.014	5653*	0.0705
Germany	-0.023**	0.046	2286**	0.03006
France	-0.0157**	0.024	2543***	0.0001
Italy	-0.028***	0.000	1168***	0.0066
Netherlands	-0.011	0.454	994	0.1423

***Significant at 99%, **Significant at 95%, *Significant at 90%

Table 2: Comparing IS medians for different samples

	Median (Pre-Covid)		Median (Post-Covid)	
	Initial sample	Average 100 simulations	Initial sample	Average 100 simulations
Spain	0	0	-0.016	-0.015
Germany	-0.01	-0.01	-0.031	-0.031
France	0	0	-0.015	-0.015
Italy	-0.002	-0.003	-0.029	-0.029
Netherlands	-0.045	-0.042	-0.059	-0.059

Columns 1 and 3 (initial sample) show IS medians for each period and country (according to the results presented in Table 3 (a) – Annex and Graph 4). Columns 2 and 4 contains the average of the 100 alternative sample.

Table 3: Results Wilcoxon Rank Sum Test for each iteration

	p-value	
	Initial sample	Average 10,000 simulations
Spain	0.060	0.08
Germany	0.065	0.073
France	0.000	0.000
Italy	0.013	0.017
Netherlands	0.220	0.247

Column 1 shows the p-values of the Wilcoxon test using the initial sample (see also table 1 in Annex D). The second column also includes the average of the p-values for each of the 10,000 combinations. See also all p-values obtained for each combination in Figure E2 in Annex E.

Table 4: Results (p-values) of the Wilcoxon Rank Sum Test for different type of reports

Type of reports	Spain	Italy	France	Germany	Netherlands
All	0.0705*	0.0066***	0.000***	0.0226**	0.142
Earnings release	0.0028***	0.0136***	0.0012***	0.0186**	0.2528

***Significant at 99%, **Significant at 95%, *Significant at 90%
p-values obtained from a Wilcoxon left-side test.

Table 5: Results (p-values) of the Wilcoxon Rank Sum Test for different dictionaries

Dictionary	Spain	Italy	France	Germany	Netherlands
LM	0.0604*	0.013**	0.0001***	0.0651*	0.220
HE	0.0015***	0.0766*	0.000***	0.179	0.304
LM modifiers	0.0705*	0.0066***	0.000***	0.0226**	0.142

***Significant at 99%, **Significant at 95%, *Significant at 90%

p-values obtained from a Wilcoxon left-side test. The dictionary "LM modifiers" is based on the LM wordlist but we adjust the sentiment if the word appears next to a modifier.

Table 6: Quantile regressions with quantitative indicators as control variables

	(1) all reports	(2) all reports	(3) earnings release	(4) earnings release
Time dummy	-0.0222***	-0.0127**	-0.0294***	-0.0140*
	(0)	(0.003)	(0)	(0.012)
EPS		0.0166**		0.0230***
		(0.002)		(0)
Constant	0	-0.00750*	0	-0.0129**
	(1)	(0.029)	(1)	(0.002)
#Observations	627	524	333	322

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns 1 to 2 regressions are based on all reports while columns 3 to 4 only refer to documents related to earnings release. The following specifications have been estimated:

$\boldsymbol{IS(\theta)_{itk} = \beta_{0\theta} + \beta_{1\theta} EPS_{it} + \delta_{t\theta} + u_{it\theta}}$ *where the subscript i refers to each bank, t to time (t=1 pre-Covid, t=2 post-Covid), k to each analyst report, and* θ^{th} *to the quantile of IS.* $\boldsymbol{\delta_{t\theta}}$ *is a dummy which takes the value 0 for the pre-Covid and 1 in the post-Covid,* $\boldsymbol{EPS_{it}}$ *is the value of released EPS of each entity, and* $u_{it\theta}$ *is the error term. Given we are evaluating each quantile of the IS values, we estimate different regressions and coefficients per each* $\boldsymbol{\theta}$*. In the given table, we evaluate the effects of each regressor in the median change of the IS, so that* θ*=0.5. Figure 12 contains different percentiles. The number of observations in the first column includes all sample reports. In the first and second column, we included all reports. Third and fourth columns restrict the analysis to earnings release documents.*

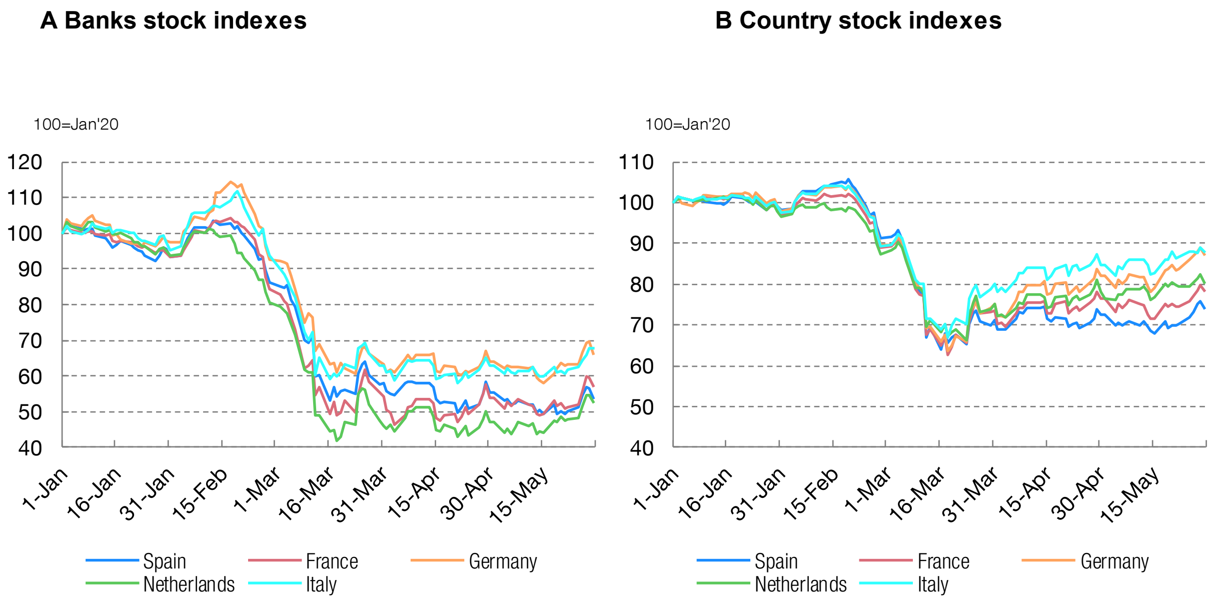
Table 7: Bank Fixed effect regressions

	(1)	(2)	(3)
EPS	0.0337^{***}		
	0		
EPS estimates		0.0107^{**}	
		(0.004)	
ROE estimates			0.00307^{**}
			(0.002)
Constant	-0.0215^{***}	-0.0251^{***}	-0.0287^{***}
	(0)	(0)	(0)
#Observations	30	30	30
R^2	0.715	0.481	0.518
RHO	0.841	0.813	0.780

p-values in parentheses

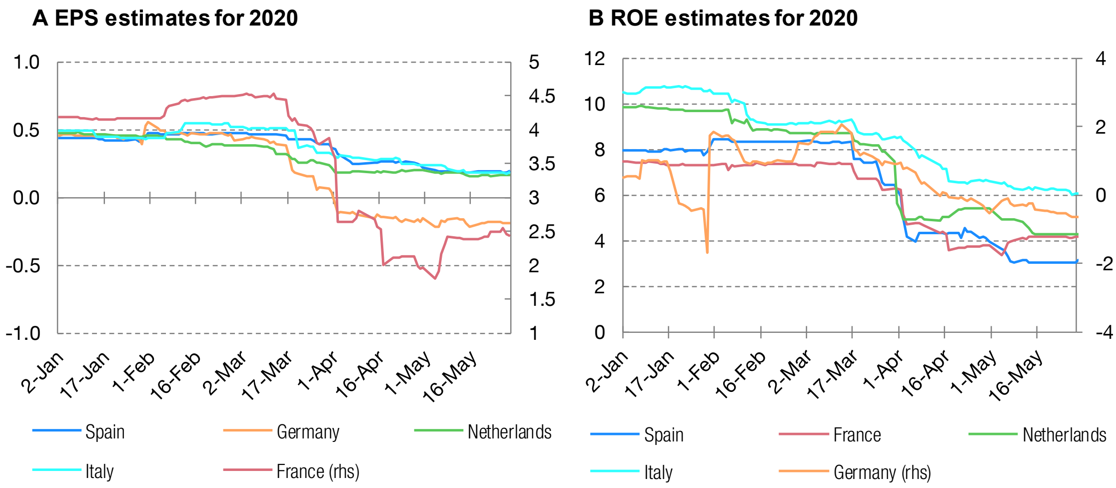
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. RHO refers to the "Intra-class" correlation and indicates how much of the total variance is explained by the differences across banks. The following specifications have been estimated: $IS_{it} = \beta_0+\beta_1 X_{it} + \alpha_i + u_{it}$, where IS_{it} refers to the median of the IS for each bank and time period t ∈{t=1 pre-Covid, t=2 post-Covid}, X_{it} shows the EPS, EPS estimates and ROE estimates, respectively, α_i is the individual bank effect not changing over time, and u_{it} is the error term. The number of observations refers to a panel containing 15 banks and 2 periods.

Figure 1: Stock market indexes



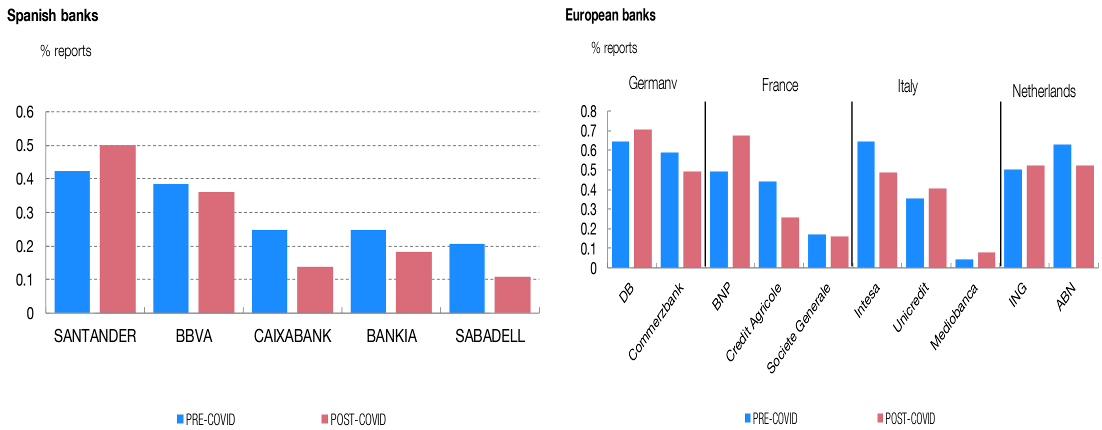
Source: Stock indexes for the banking system in each country (Datastream)

Figure 2: Analysts estimates (2020) for European banks



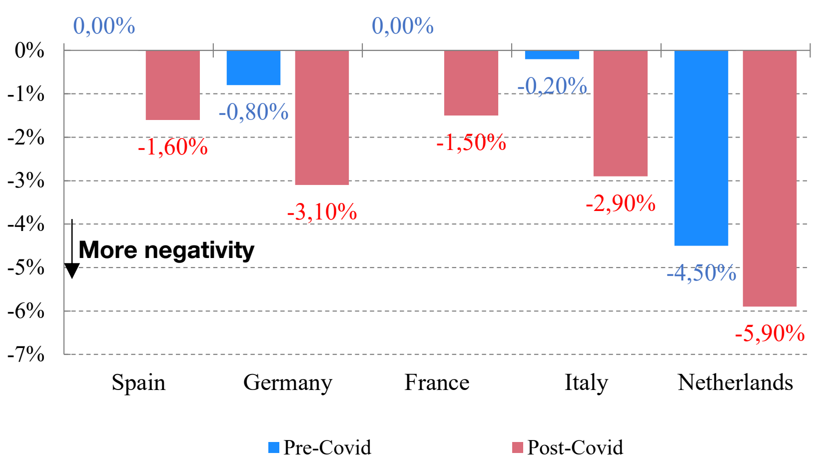
We considered the SmartEstimate, which is a weighted average of analysts' estimates provided less than 120 days before. Two-thirds of the weighting is obtained from contributor punctuation and one third depends on the seniority of the estimation. The evolution in each country is constructed through a weighted average of the stock market capitalization of each bank.

Figure 3: Percentage of reports where each bank is mentioned



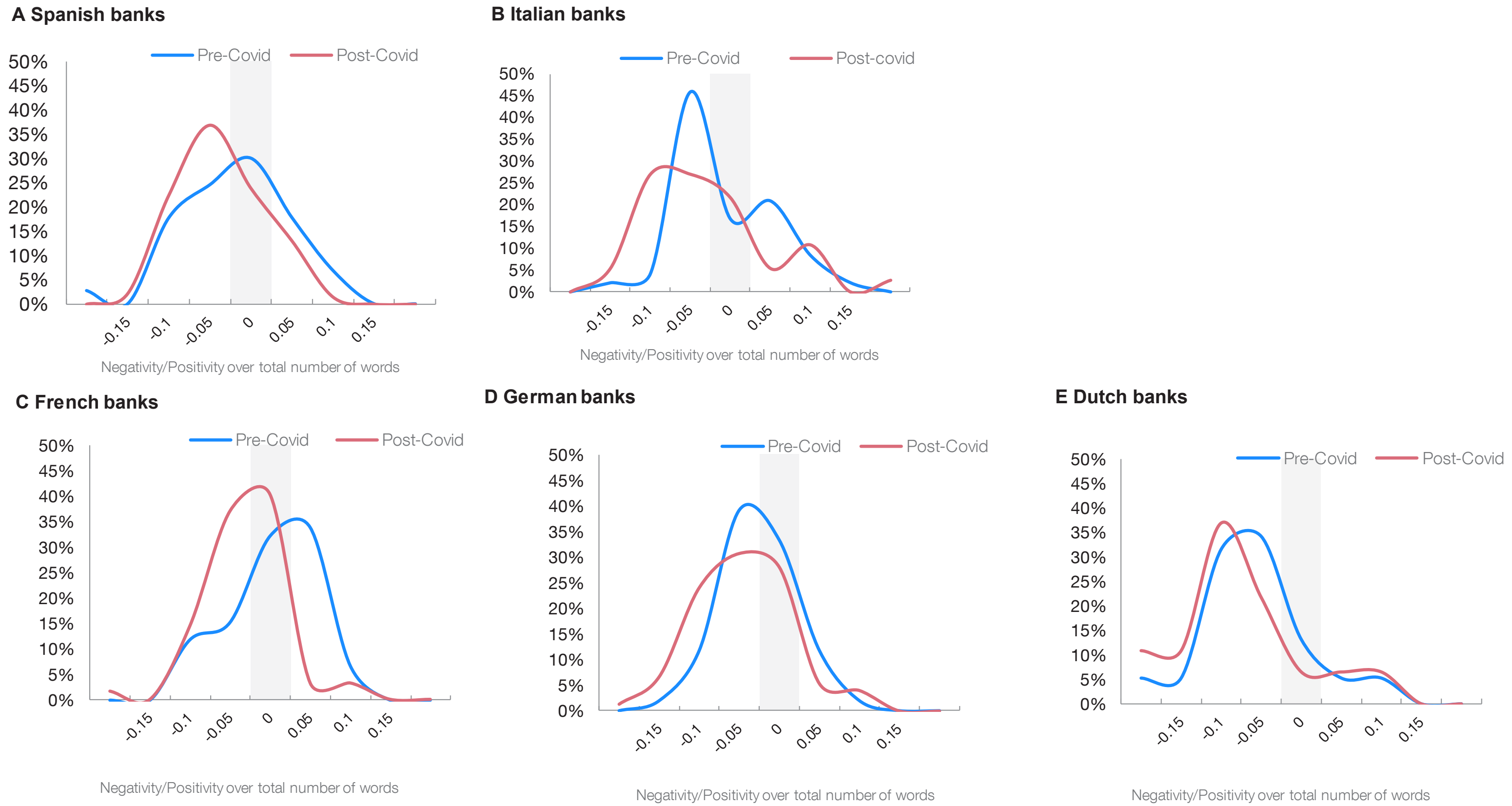
Note: the sum of the percentages can be bigger than one, as some of the reports refer to more than one entity.

Figure 4: Sentiment Index before and after Covid



Median of the Sentiment Index in each country and period. The index is expressed as a percentage of total number of words.

Figure 5: Analysts' sentiment distribution for European banks



Figures show the percentage of reports in each range of IS values. The vertical bar highlights the reports having a neutral sentiment, i.e., the ones neither with positive nor negative words or the ones showing equivalence between the number of positive and negative words. Observations in the left-side show negative values while in the right-side one can see the positive values.

Figure 6: Percentage of positivity and negativity in the Sentiment Index

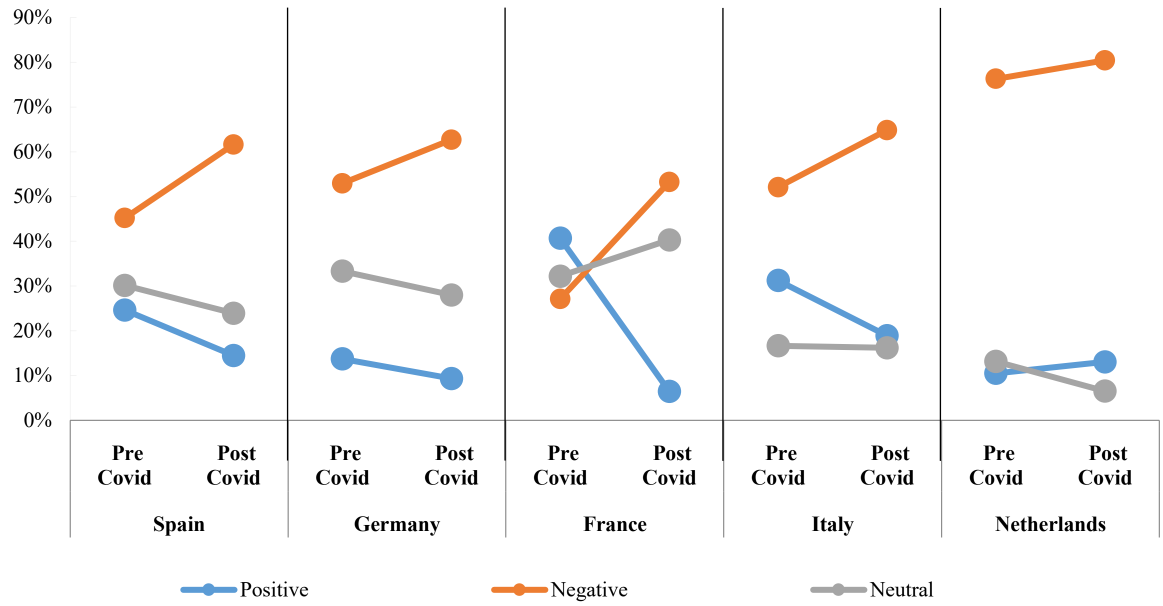


Figure shows the percentage of reports having a positive, negative or neutral sentiment.

Figure 7: The impact of Covid on Analysts's opinions about the banking sector (difference in medians)

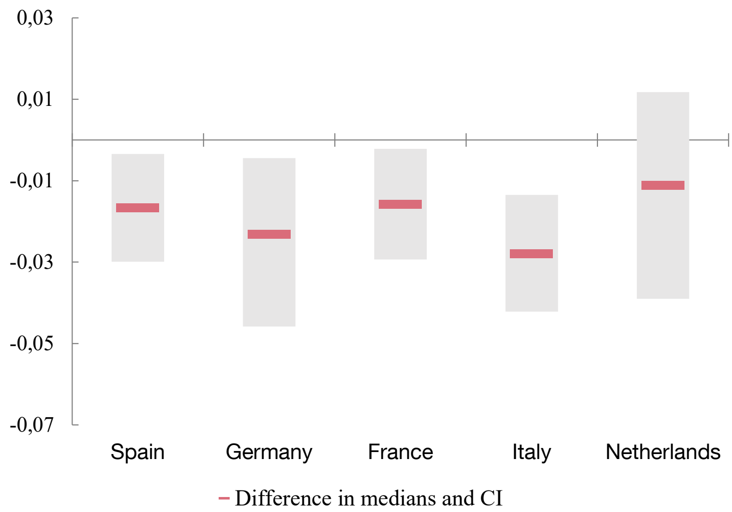
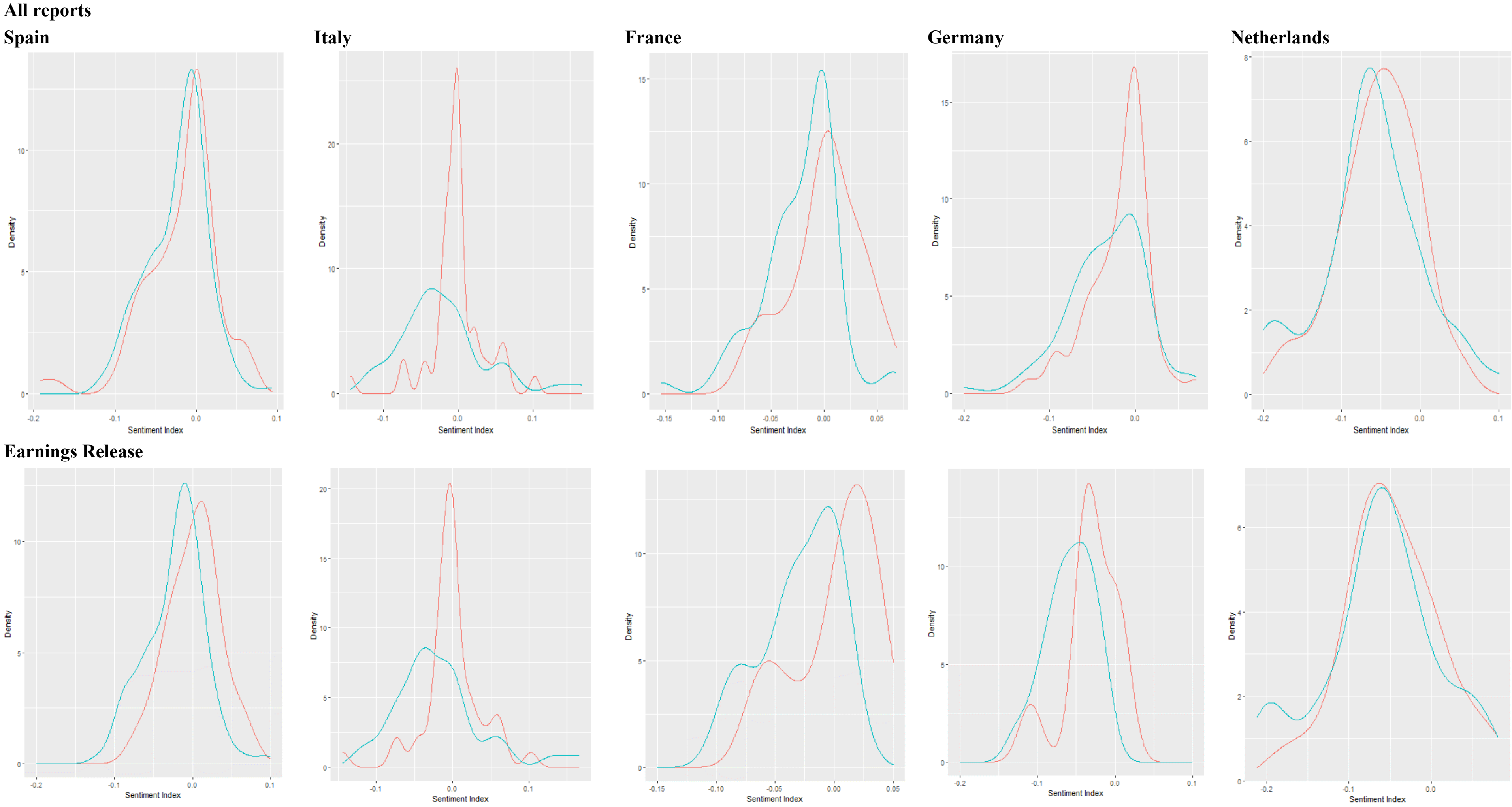


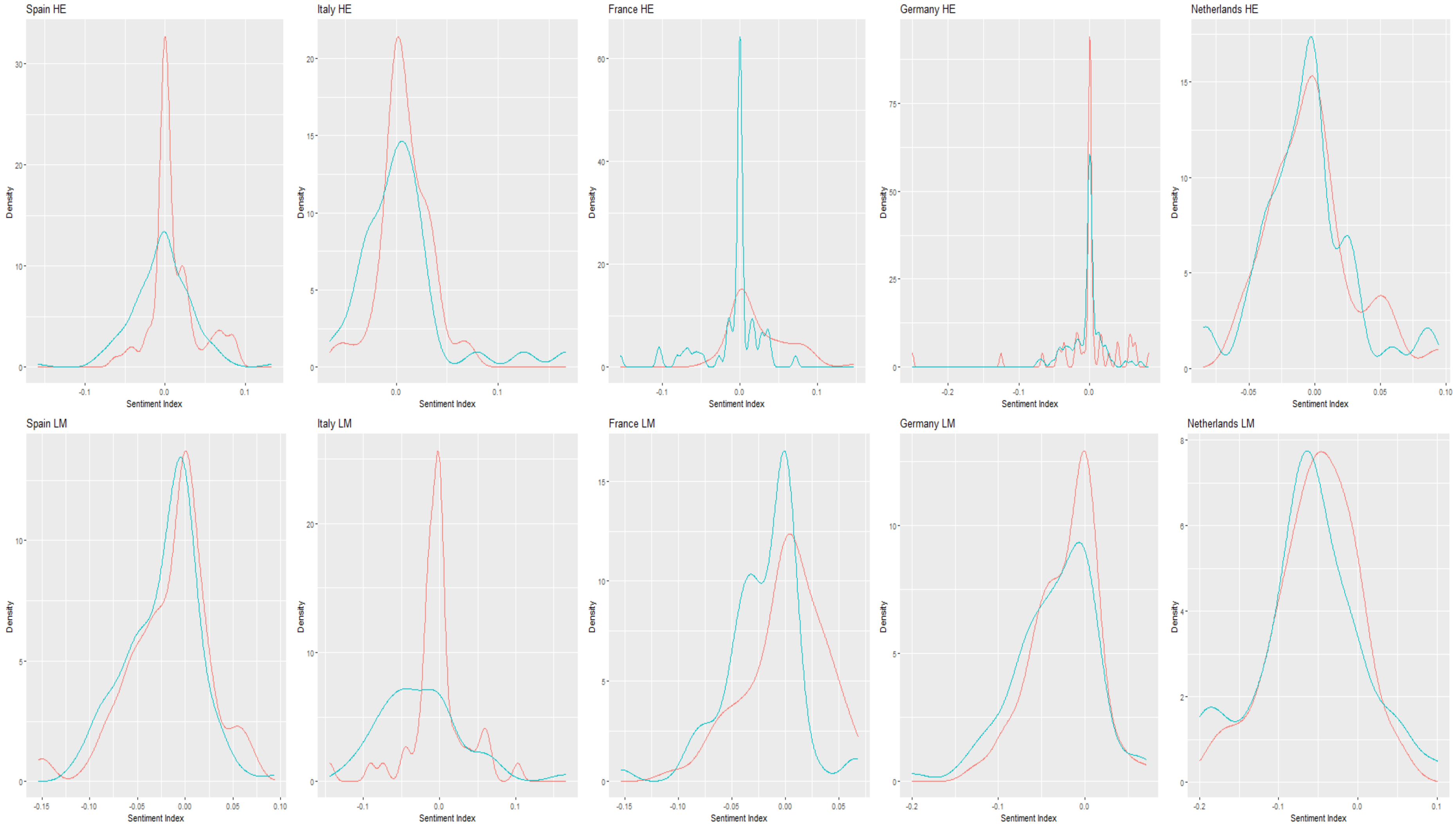
Figure shows the differences between post-Covid and pre-Covid median and a 95% CI.

Figure 8: Sentiment Index Distributions for Earnings Release and all the reports



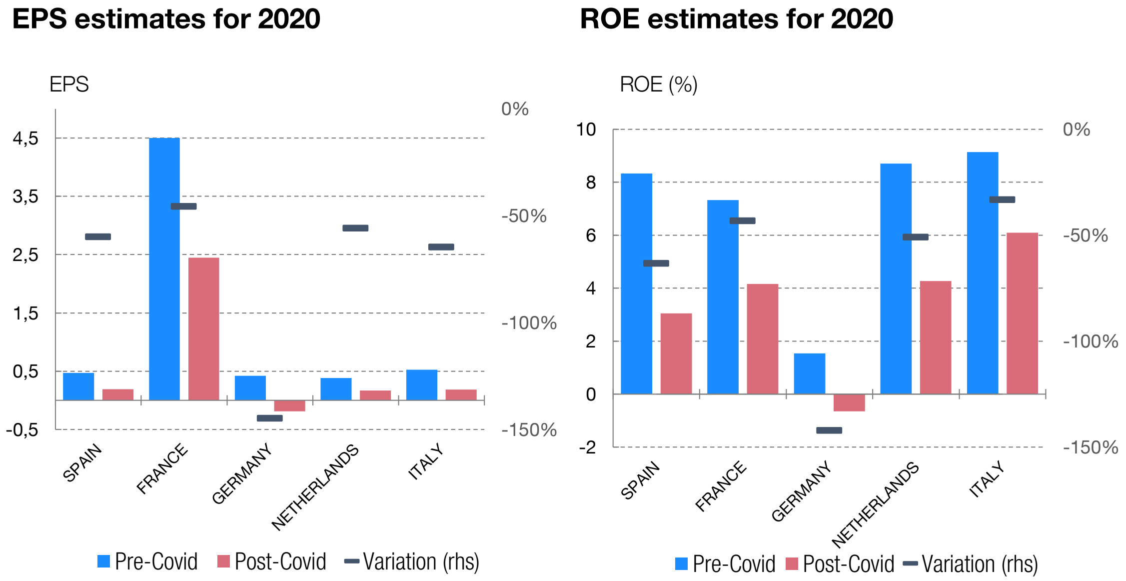
Red lines represent the pre-covid densities, and the blue lines the post-covid ones.

Figure 9: Comparison between HE and LM



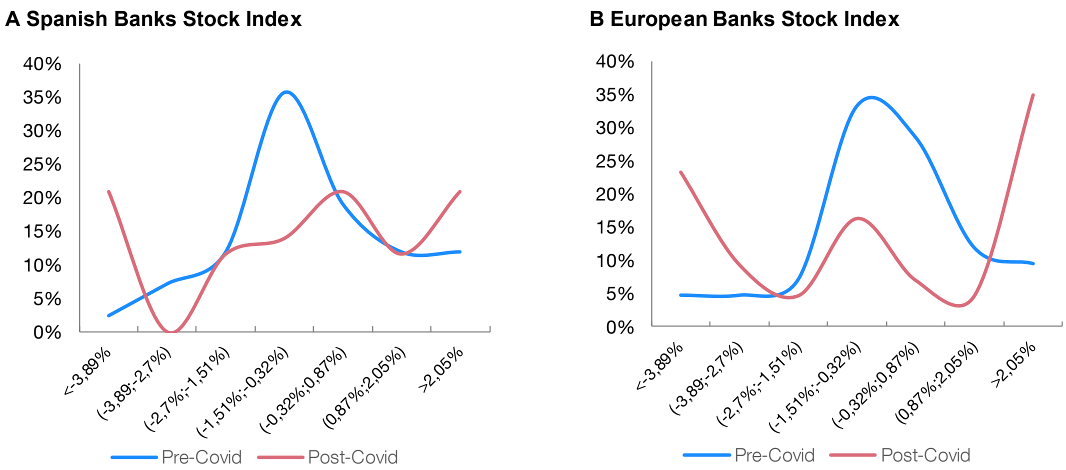
Red lines represent the pre-covid densities, and the blue lines the post-covid ones.

Figure 10: Analysts' estimates before and after Covid pandemic



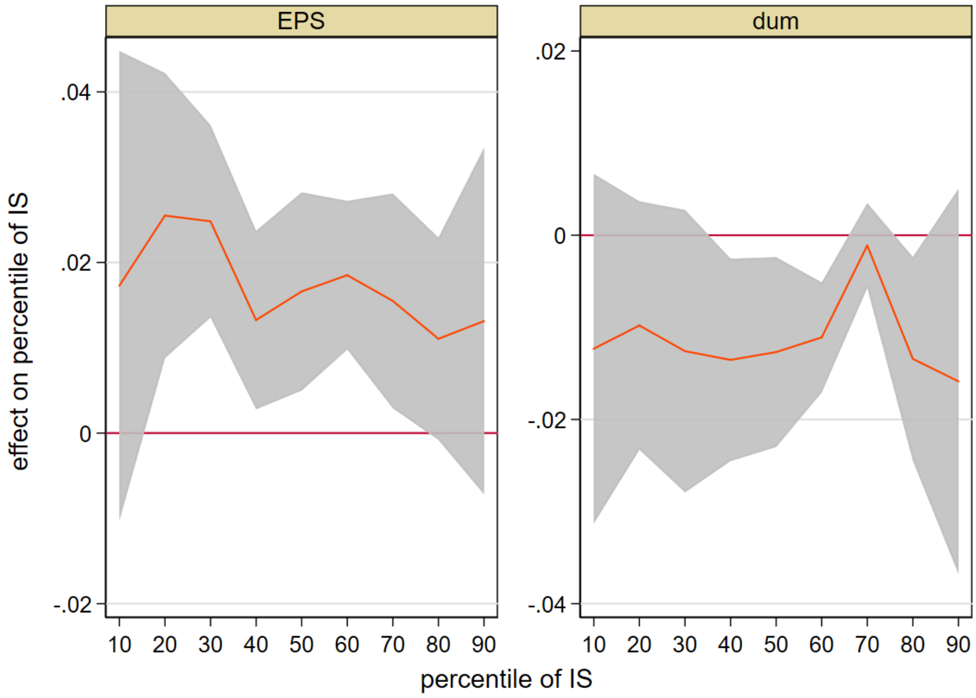
We considered the SmartEstimate, which is a weighted average of analysts' estimates provided less than 120 days before. Two-thirds of the weighting is obtained from contributor punctuation and one third depends on the seniority of the estimation. The evolution in each country is constructed through a weighted average of the stock market capitalization of each bank. The pre-Covid value represents last February value and the post-Covid the last data on May.

Figure 11: Stock daily returns histograms before and after Covid



Daily returns distributions for the Spanish and european banks stock index during January and February (pre-Covid) and May, April (post-Covid).

Figure 12: Quantile regression plots for the effect of each variable in the percentiles of IS



Dum: dummy variable for time (0: pre-Covid, 1: post-Covid).

Annex A: Summary statistics for the sample of reports

Table A1: Sample size

	Number of reports			% Market Capitalization sample
	Pre-Covid	Post-Covid	Total	
Spain	73	138	211	93%
Germany	51	75	126	80%
France	59	62	121	97%
Italy	48	37	85	70%
Netherlands	38	46	84	90%
Total	**269**	**358**	**627**	

Table A2: Number of Banks (percentage) included in each report

	Spanish banks		European banks	
# Banks	**PRE-COVID**	**POST-COVID**	**PRE-COVID**	**POST-COVID**
1	66%	83%	87%	85%
2	23%	9%	13%	11%
3	0%	1%	0%	1%
4	3%	1%	-	-
5	5%	4%	-	-
General	3%	1%	0%	0%

Table A3: Median and standard deviation for Sentiment Index (IS)

a. The ratio over the total number of words

Country	Median		Standard deviation	
	Pre-Covid	Post-Covid	Pre-Covid	Post-Covid
Spain	0	-0.016	0.0461	0.0368
Germany	-0.008	-0.031	0.0305	0.0463
France	0	-0.015	0.0349	0.0366
Italy	-0.002	-0.029	0.0382	0.0612
Netherlands	-0.045	-0.059	0.0544	0.0659

b. The ratio over the terms

Country	Median		Standard deviation	
	Pre-Covid	Post-Covid	Pre-Covid	Post-Covid
Spain	0.00	-0.33	0.5717	0.5498
Germany	0.00	-0.41	0.6024	0.5572
France	0.00	-0.23	0.6670	0.4902
Italy	-0.04	-0.37	0.5561	0.6086
Netherlands	-0.70	-0.72	0.4706	0.5496

Annex B: Positivity, Negativity and Sentiment index over the total of connote terms

In order to construct this alternative version of the index, we should consider only connote terms. That way, the index for each document shows the relevance of negative (positive) words over the total number of connote terms. Equations B1, B2 and B3 show the formulas to compute the indexes.

$$\text{Negativity Index} = \frac{\sum \# \, Negative\ words}{\sum \#\text{Negative words} + \sum \#\text{Positive words}} \tag{B1}$$

$$\text{Positivity Index} = \frac{\sum \# \, Positive\ words}{\sum \#\text{Negative words} + \sum \#\text{Positive words}} \tag{B2}$$

$$\text{Sentiment Index} = \frac{\sum \# \, Positive\ words - \sum \# \, Negative\ words}{\sum \#\text{Negative words} + \sum \#\text{Positive words}} \tag{B3}$$

Annex C: Correlation between different indexes (Spanish banks)

	Pre-covid	Post-covid
Model 1-2	-0.856	-0.844

Model 1: Sentiment index computed using the total number of words, Model 2: Sentimen index based on connote terms.

Annex D: Hypothesis statistical test

Several hypothesis tests have been used to assess the significance of sentiment changes before and after the inception of the pandemic. The Kolmogorov-Smirnov, the Chi-squared, and the Wilcoxon Rank Sum Test can be used to compare the distributions of the Sentiment index in the two periods. Additionally, a Quantile Regression can determine whether if there is a change in median's difference in the two periods and the Fligner-Kileen test provides a statistical tool to analyse if standard deviations of the two distributions changed.

Our results rely mainly on the Wilcoxon Test, the Quantile Regression, and the Fligner-Kileen test. The Fligner-Kileen constitutes a non-parametric test that can be used to assess opinion consensus among analyst's sentiment. The Wilcoxon-test is the non-parametric extension of the t-test and assesses if two samples are obtained from homogeneous populations, i.e., if there are the same number of positive and negative differences across the samples and if the magnitude of the differences is the same (symmetry of positive and negative differences). Using quantile regression, we test for differences in medians.

However, the Chi-squared and the Kolmogorov-Smirnov tests should not be as appropriate in this context. In the first case, the Chi-squared requires the two samples to be independent, which is difficult to guarantee in our sample of reports. In the second case, the Kolmogorov-Smirnov test is not able to identify locations shifts as its main aim is looking for shape changes. Each of the tests is defined as follows:

1. **Chi-squared**: is a non-parametric test that compares observed vs expected frequencies. It does not require equal variance among the samples but it required the two analysed groups to be independent, i.e., the persistence in time series data

can affect the robustness of the statistic. Moreover, the sample size should be large enough.

2. **Quantile regression**: allows to statistically test whether there was a change in median sentiment after Covid. For this purpose, quantile regression is estimated where the independent variable is the sentiment index and the explanatory variable is a dichotomous variable that identifies the pre-covid and post-covid periods (0=Ener-Feb, 1=Apr-May). This model does not assume that the data must follow a specific type of distribution (so it is considered a semi-parametric test). The null hypothesis assumes equality at the median.

3. **Kolmogorov test:** A nonparametric test based on the cumulative density function that tests the similarity between two distributions, mainly in their shape. However, it has the disadvantage of being less powerful in detecting changes in the location of the median than other tests. The null hypothesis assumes equality of distributions.

4. **Wilcoxon rank-sum test:** A non-parametric test, which does not impose any kind of functional form. It is, therefore, the extension of the t-test when the samples do not follow a normal distribution. It assumes independence between samples and equality of variances. The null hypothesis is equality of medians and same location of the distribution. The alternative hypothesis would indicate that the medians of the two distributions are different and/or that one of the distributions has higher (or lower) values than the other.

5. **Fligner-Killeen test**: non-parametric test to evaluate the equality of variance between groups. Mainly powerful when distributions are not normal and there are outliers. The null hypothesis assumes equality of variance.

Table D1: Comparison of results for different tests

	Test Chi-squared			Quantile Regression			Test Kolmogorov			Fligner-Killeen Test			Wilcoxon Rank Sum Test (a)		
	Test		*p*-value	Test		*p*-value	Test		*p*-value	Test		*p*-value	Test		*p*-value
Spain	2600	***	0.005	-0.016	**	0.035	-0.143		0.142	0.078		0.780	6259	*	0.060
Germany	935		0.427	-0.022	*	0.065	-0.163		0.199	1.810		0.179	2213	*	0.065
France	1400	***	0.002	-0.016	**	0.018	-0.342	***	0.001	0.080		0.775	2516	***	0.000
Italy	828		0.525	-0.031	***	0.000	-0.415	***	0.001	8.279	***	0.004	1138	**	0.013
Netherlands	1203		0.120	-0.011		0.444	-0.205		0.175	0.697		0.404	960		0.220

**** Significant at 99%, ** Significant at 95%, * Significant at 90%*

(a) The alternative hypothesis that the post-Covid distribution shows a higher proportion of values to the left (negative values) than the pre-Covid distribution.

Table D2: Quantile regression for the two methodologies to compute sentiment index

	Quantile Regression					
	ratio over the total number of words			ratio over connote terms		
	Test		*p*-value	Test		*p*-value
Spain	-0.016	**	0.035	-0.273	**	0.020
Germany	-0.022	*	0.065	-0.207		0.348
France	-0.016	**	0.018	-0.255	**	0.048
Italy	-0.031	***	0.000	-0.196	*	0.078
Netherlands	-0.011		0.444	0.026		0.858

**** Significant at 99%, ** Significant at 95%, * Significant at 90%*

ANNEX E: Robustness analysis

Figure E1: Sentiment Index distributions for different samples

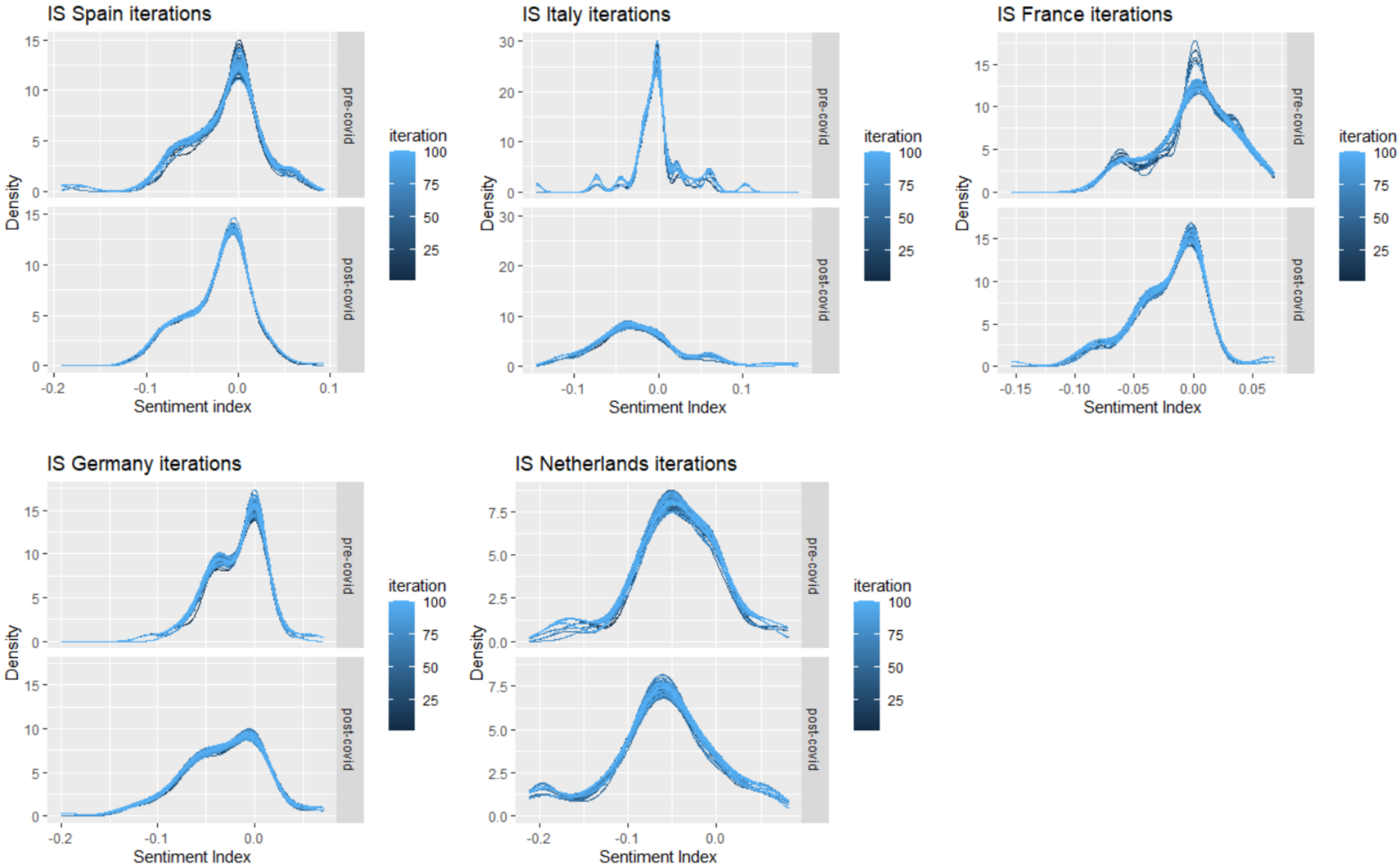
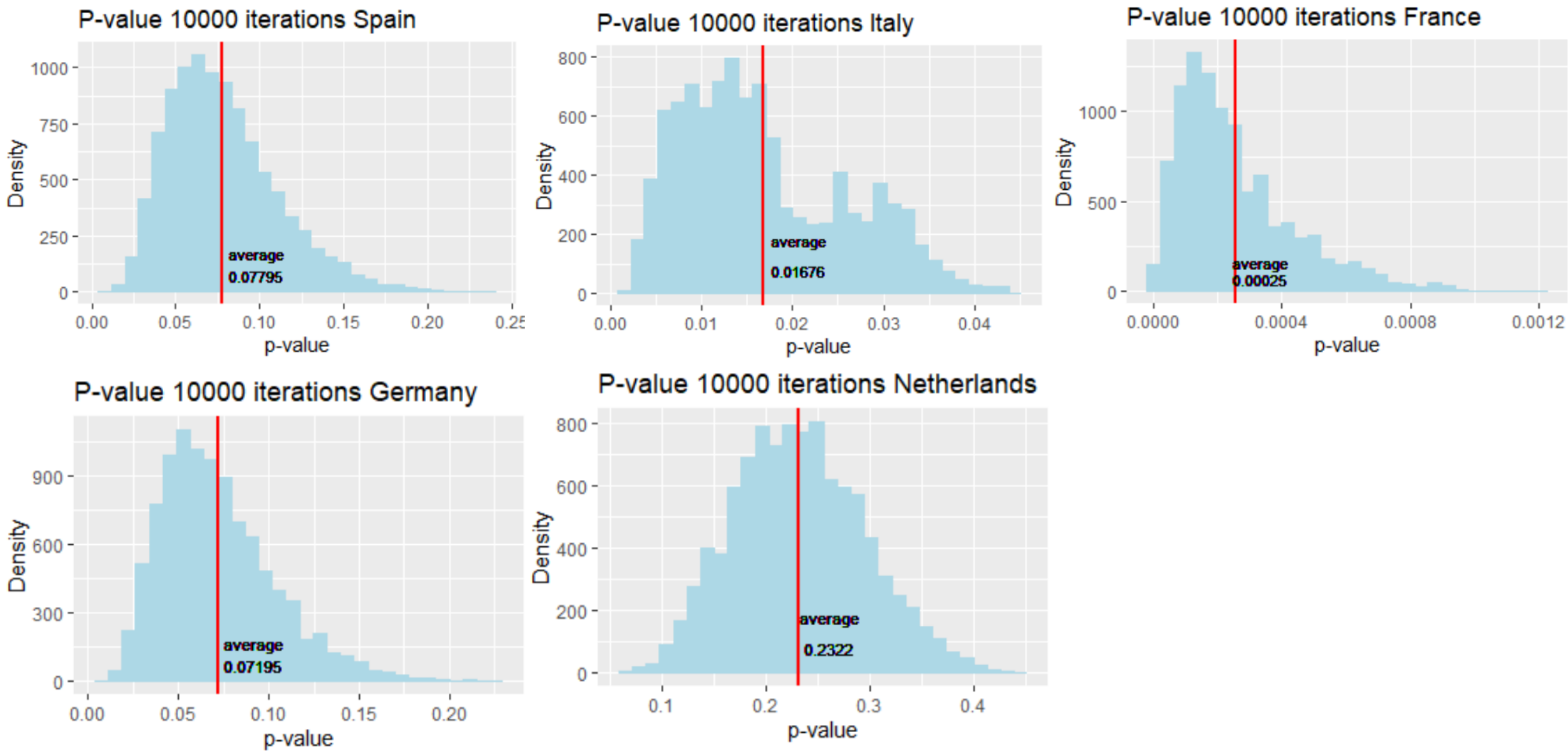


Figure E2: P-values for each sample iteration (Wilcoxon Rank Sum Test)



Source: own elaboration, the chart shows the p-value distribution for each sample combinations and the mean.

ANNEX F: Positive and negative words in HE Dictionary

Negative words		Positive Words		Modifiers	
below	obstacles	above	improvements	ain't	needn't
challenge	penalties	accomplish	improves	aint	needn’t
challenged	penalty	accomplished	improving	aren't	Neither
challenges	risk	accomplishes	increase	arent	Never
challenging	risks	accomplishing	increased	can't	No
decline	risky	accomplishment	increases	cannot	Nobody
declined	shrink	accomplishments	increasing	cant	None
declines	shrinking	achieve	larger	couldn't	Nor
declining	shrinks	achieved	largest	couldnt	Not
decrease	shrunk	achievement	leader	daren't	oughtn't
decreased	slump	achievements	leading	darent	oughtn’t
decreases	slumped	achieves	leading	didn't	shan't
decreasing	slumping	achieving	more	didnt	Shant
depressed	slumps	beat	most	doesn't	shouldn't
deteriorate	smaller	beating	ncouraging	doesnt	shouldn’t
deteriorated	smallest	beats	opportunities	don't	wasn't
deteriorates	threat	best	opportunity	dont	wasn’t
deteriorating	threats	better	pleased	hadn't	weren't
difficult	uncertain	certain	positive	hadnt	weren’t
difficulty	uncertainty	certainty	positives	hasn't	won't
disappoint	under	definite	progress	hasnt	Wont
disappointed	unfavorable	deliver	progressing	haven't	wouldn't
disappointing	unsettled	delivered	record	havent	wouldn’t
disappointment	weak	delivering	reward	isn't	
disappoints	weaken	delivers	rewarded	isnt	
down	weakened	encouraged	rewarding	mightn't	
downturn	weakening	enjoy	rewards	mightnt	
drop	weakens	enjoyed	rise	mustn't	
dropped	weakness	enjoying	risen	mustnt	
dropping	weaknesses	enjoys	rises		
drops	worse	exceed	rising		
fail	worsen	exceeded	rose		
failing	worsening	exceeding	Solid		
fails	worsens	exceeds	Strength		
failure	worst	excellent	Strengthen		
fall		expand	Strengthened		
fallen		expanded	Strengthening		
falling		expanding	Strengthens		
falls		expands	Strengths		
fell		expansion	Strong		
hurdle		good	Stronger		
hurdles		greater	Strongest		
least		greatest	Succeed		
less		grew	Succeeded		
low		grow	Succeeding		
lower		growing	Succeeds		
lowest		grown	Success		
negative		grows	Successes		
negatives		growth	Successful		
obstacle		high	Up		
		higher			
		highest			
		improve			
		improved			
		improvement			

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2010 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: Eurozone prices: a tale of convergence and divergence.

2011 ÁNGEL IVÁN MORENO BERNAL and CARLOS GONZÁLEZ PEDRAZ: Sentiment analysis of the Spanish Financial Stability Report. (There is a Spanish version of this edition with the same number).

2012 MARIAM CAMARERO, MARÍA DOLORES GADEA-RIVAS, ANA GÓMEZ-LOSCOS and CECILIO TAMARIT: External imbalances and recoveries.

2013 JESÚS FERNÁNDEZ-VILLAVERDE, SAMUEL HURTADO and GALO NUÑO: Financial frictions and the wealth distribution.

2014 RODRIGO BARBONE GONZALEZ, DMITRY KHAMETSHIN, JOSÉ-LUIS PEYDRÓ and ANDREA POLO: Hedger of last resort: evidence from Brazilian FX interventions, local credit, and global financial cycles.

2015 DANILO LEIVA-LEON, GABRIEL PEREZ-QUIROS and EYNO ROTS: Real-time weakness of the global economy: a first assessment of the coronavirus crisis.

2016 JAVIER ANDRÉS, ÓSCAR ARCE, JESÚS FERNÁNDEZ-VILLAVERDE and SAMUEL HURTADO: Deciphering the macroeconomic effects of internal devaluations in a monetary union.

2017 JACOPO TIMINI, NICOLA CORTINOVIS and FERNANDO LÓPEZ VICENTE: The heterogeneous effects of trade agreements with labor provisions.

2018 EDDIE GERBA and DANILO LEIVA-LEON: Macro-financial interactions in a changing world.

2019 JAIME MARTÍNEZ-MARTÍN and ELENA RUSTICELLI: Keeping track of global trade in real time.

2020 VICTORIA IVASHINA, LUC LAEVEN and ENRIQUE MORAL-BENITO: Loan types and the bank lending channel.

2021 SERGIO MAYORDOMO, NICOLA PAVANINI and EMANUELE TARANTINO: The impact of alternative forms of bank consolidation on credit supply and financial stability.

2022 ALEX ARMAND, PEDRO CARNEIRO, FEDERICO TAGLIATI and YIMING XIA: Can subsidized employment tackle long-term unemployment? Experimental evidence from North Macedonia.

2023 JACOPO TIMINI and FRANCESCA VIANI: A highway across the Atlantic? Trade and welfare effects of the EU-Mercosur agreement.

2024 CORINNA GHIRELLI, JAVIER J. PÉREZ and ALBERTO URTASUN: Economic policy uncertainty in Latin America: measurement using Spanish newspapers and economic spillovers.

2025 MAR DELGADO-TÉLLEZ, ESTHER GORDO, IVÁN KATARYNIUK and JAVIER J. PÉREZ: The decline in public investment: "social dominance" or too-rigid fiscal rules?

2026 ELVIRA PRADES-ILLANES and PATROCINIO TELLO-CASAS: Spanish regions in Global Value Chains: How important? How different?

2027 PABLO AGUILAR, CORINNA GHIRELLI, MATÍAS PACCE and ALBERTO URTASUN: Can news help measure economic sentiment? An application in COVID-19 times.

2028 EDUARDO GUTIÉRREZ, ENRIQUE MORAL-BENITO, DANIEL OTO-PERALÍAS and ROBERTO RAMOS: The spatial distribution of population in Spain: an anomaly in European perspective.

2029 PABLO BURRIEL, CRISTINA CHECHERITA-WESTPHAL, PASCAL JACQUINOT, MATTHIAS SCHÖN and NIKOLAI STÄHLER: Economic consequences of high public debt: evidence from three large scale DSGE models.

2030 BEATRIZ GONZÁLEZ: Macroeconomics, Firm Dynamics and IPOs.

2031 BRINDUSA ANGHEL, NÚRIA RODRÍGUEZ-PLANAS and ANNA SANZ-DE-GALDEANO: Gender Equality and the Math Gender Gap.

2032 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Machine learning in credit risk: measuring the dilemma between prediction and supervisory cost.

2033 PILAR GARCÍA-PEREA, AITOR LACUESTA and PAU ROLDAN-BLANCO: Raising Markups to Survive: Small Spanish Firms during the Great Recession.

2034 MÁXIMO CAMACHO, MATÍAS PACCE and GABRIEL PÉREZ-QUIRÓS: Spillover Effects in International Business Cycles.

2035 ÁNGEL IVÁN MORENO and TERESA CAMINERO: Application of text mining to the analysis of climate-related disclosures.

2036 EFFROSYNI ADAMOPOULOU and ERNESTO VILLANUEVA: Wage determination and the bite of collective contracts in Italy and Spain: evidence from the metal working industry.

2037 MIKEL BEDAYO, GABRIEL JIMÉNEZ, JOSÉ-LUIS PEYDRÓ and RAQUEL VEGAS: Screening and Loan Origination Time: Lending Standards, Loan Defaults and Bank Failures.

2038 BRINDUSA ANGHEL, PILAR CUADRADO and FEDERICO TAGLIATI: Why cognitive test scores of Spanish adults are so low? The role of schooling and socioeconomic background

2039 CHRISTOPH ALBERT, ANDREA CAGGESE and BEATRIZ GONZÁLEZ: The Short- and Long-run Employment Impact of COVID-19 through the Effects of Real and Financial Shocks on New Firms.

2040 GABRIEL JIMÉNEZ, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Who Truly Bears (Bank) Taxes? Evidence from Only Shifting Statutory Incidence.

2041 FELIX HOLUB, LAURA HOSPIDO and ULRICH J. WAGNER: Urban air pollution and sick leaves: evidence from social security data.

2042 NÉLIDA DÍAZ SOBRINO, CORINNA GHIRELLI, SAMUEL HURTADO, JAVIER J. PÉREZ and ALBERTO URTASUN: The narrative about the economy as a shadow forecast: an analysis using Banco de España quarterly reports.

2043 NEZIH GUNER, JAVIER LÓPEZ-SEGOVIA and ROBERTO RAMOS: Reforming the individual income tax in Spain.

2101 DARÍO SERRANO-PUENTE: Optimal progressivity of personal income tax: a general equilibrium evaluation for Spain.

2102 SANDRA GARCÍA-URIBE, HANNES MUELLER and CARLOS SANZ: Economic uncertainty and divisive politics: evidence from the *Dos Españas*.

2103 IVÁN KATARYNIUK, VÍCTOR MORA-BAJÉN and JAVIER J. PÉREZ: EMU deepening and sovereign debt spreads: using political space to achieve policy space.

2104 DARÍO SERRANO-PUENTE: Are we moving towards an energy-efficient low-carbon economy? An input-output LMDI decomposition of CO_2 emissions for Spain and the EU28.

2105 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation.

2106 JAVIER ANDRÉS, ÓSCAR ARCE and PABLO BURRIEL: Market polarization and the Phillips curve.

2107 JUAN de LUCIO and JUAN S. MORA-SANGUINETTI: New dimensions of regulatory complexity and their economic cost. An analysis using text mining.

2108 DANILO LEIVA-LEON and LUIS UZEDA: Endogenous time variation in vector autoregressions.

2109 CORINNA GHIRELLI, ANDREA GONZÁLEZ, JOSÉ LUIS HERRERA and SAMUEL HURTADO: Weather, mobility and the evolution of the Covid-19 pandemic.

2110 KLODIANA ISTREFI, FLORENS ODENDAHL and GIULIA SESTIERI: Fed communication on financial stability concerns and monetary policy decisions: revelations from speeches.

2111 YUNUS AKSOY, HENRIQUE S. BASSO and CAROLYN ST AUBYN: Time Variation in Lifecycle Consumption and Income.

2112 JENNIFER PEÑA and ELVIRA PRADES: Price setting in Chile: micro evidence from consumer on-line prices during the social outbreak and Covid-19.

2113 NEZIH GUNER, YULIYA A. KULIKOVA and ARNAU VALLADARES-ESTEBAN: Does the added worker effect matter?

2114 RODOLFO G. CAMPOS and JACOPO TIMINI: Unequal trade, unequal gains: the heterogeneous impact of MERCOSUR.

2115 JAVIER QUINTANA: Import competition, regional divergence, and the rise of the skilled city.

2116 PATRICK MACNAMARA, MYROSLAV PIDKUYKO and RAFFAELE ROSSI: Marginal Tax Changes with Risky Investment.

2117 RODOLFO G. CAMPOS, JACOPO TIMINI and ELENA VIDAL: Structural gravity and trade agreements: does the measurement of domestic trade matter?

2118 ALESSANDRO MELCARNE, JUAN S. MORA-SANGUINETTI and ROK SPRUK: Democracy, technocracy and economic growth: evidence from 20 century Spain.

2119 ÁNGEL ESTRADA and DANIEL SANTABÁRBARA: Recycling carbon tax revenues in Spain. Environmental and economic assessment of selected green reforms.

2120 ALEJANDRO FERNÁNDEZ-CEREZO, BEATRIZ GONZÁLEZ, MARIO IZQUIERDO and ENRIQUE MORAL-BENITO: Firm-level heterogeneity in the impact of the COVID-19 pandemic.

2121 EDUARDO GUTIÉRREZ and CÉSAR MARTÍN MACHUCA: The effect of tariffs on Spanish goods exports.

2122 JACOPO TIMINI: Revisiting the 'Cobden-Chevalier network' trade and welfare effects.

2123 ALEJANDRO BUESA, JAVIER J. PÉREZ and DANIEL SANTABÁRBARA: Awareness of pandemics and the impact of COVID-19.

2124 ALICIA AGUILAR and DIEGO TORRES: The impact of COVID-19 on analysts' sentiment about the banking sector.